DOCUMENT RESUME

ED 443 874 TM 031 519

AUTHOR	Mendro, Robert; Bembry, Karen
TITLE	School Evaluation: A Change in Perspective.
PUB DATE	2000-04-27
NOTE	22p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Accountability; *Decision Making; Elementary Secondary Education; *Evaluation Methods; Program Evaluation; *School Effectiveness; Student Evaluation
IDENTIFIERS	*Dallas Independent School District TX

ABSTRACT

This paper shows the shift in emphasis in school evaluation that has occurred in the Dallas Public Schools, Texas, in recent years. The primary emphasis is now on accountability in measuring, using, and learning from teacher and school effectiveness. Both value-added and unadjusted measures are needed for these evaluations, and traditional program evaluation has been repositioned to provide light on teacher and school effectiveness whenever possible. Traditional activities are still responsible for many compliance areas, but they are increasingly focused on effectiveness information. As a concomitant of the effectiveness studies, training and service activities have taken a major place in the school system's Department of Research and Evaluation. They were once nearly perfunctory, but now they are used to expand knowledge of the accountability system and effective practices and to help administrators make better decisions. (Contains 26 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

School Evaluation: A Change in Perspective

Robert Mendro
Karen Bembry

Dallas Public Schools

R. Mendro

1

Introduction

School evaluation has changed considerably in the last decade in the Dallas Public Schools. The shift has been a move away from program evaluation as the primary purpose and activity of the Department of Research and Evaluation and a move toward determining and using school and teacher effectiveness. The change has come about because of a change in national and state focus, new statistical tools that were either largely unavailable or unknown prior to the 1990s, and because of much old and some new research and evaluation that has changed the focus of our efforts.

For many years the District and Department followed what was the classical evaluation approach of the late 1960s and early 1970s. New curricula or new programs or new approaches to presenting information were the way to go about changing the schools. Good program evaluation was the key to determining which new curricula, which new programs, or which new way of presenting information was going to move us forward.

We kept running into a consistent problem though. We measured the results of program after program and very few had noticeable positive effects and some had negative effects. (See for example the program evaluations in Dallas Public Schools, 1982 and Dallas Public Schools, 1987). With few exceptions, the programs where effects were found replicated poorly when the program was expanded. Finally, as our process evaluation revealed, many were poorly implemented. After many approaches to the problem and

ED 443 874

TM031519

Paper presented in Symposium at the 2000 Annual Meeting of the American Educational Research Association. April 27, 2000, New Orleans.

BEST COPY AVAILABLE

much solid evaluation of failing efforts, we finally were convinced of a major flaw in our expectations. Our search for the program that was relatively straightforward for the general teacher to understand and implement and that was also relatively safe from the effects of less able or interested teachers was futile.

Along the way, we were required to and were also interested in the idea of measuring effective and ineffective schools. In the 1980s we tried a comprehensive program to identify and reward effective schools (Webster and Olson, 1988). A multiple regression approach was used to determine a growth curve for each student and effective schools were identified. The intent was to eliminate the bias that comes from measuring effective achievement when achievement is correlated with so many factors outside the control of the school. Our research into the system showed that indeed, we controlled bias at the student level. Results for students were mostly uncorrelated with SES, ethnicity, or language proficiency. However, we found that the results were still significantly correlated (in the practical sense, our numbers were always large enough to guarantee statistical significance) with school characteristics (Webster and Olson, 1988). When the state established a career ladder program that absorbed the funds for the reward program, we discontinued the effort, in part because of the then intractable problem of school level correlations.

In 1992, following the report of a special Board of Education Commission, the District began measuring school effectiveness using a different multiple regression approach (Webster, Mendro, and Almaguer, 1993; Commission for Educational Excellence, 1991). As a part of this approach, school level variables were partially controlled by using them as variables in the student level equations. We began research on models of school

effectiveness and began looking into multi-level modeling, in particular hierarchical linear modeling (HLM). In 1993-94, the Board mandated that we measure teacher effectiveness and in 1995, we began using HLM in the computation of our effectiveness measures. (Mendro, Webster, Bembry, and Orsak, 1995; Webster 1995; Webster and Mendro, 1997; Webster, Mendro, Bembry, and Orsak, 1995.)

From 1994 onward, we began a series of evaluations and research papers using school effectiveness data and from 1997 conducting research using teacher effectiveness data (Bearden, 1997; Jordan, Mendro, and Weerasinghe, 1997; Webster, Mendro, Bearden, Bembry, and Jordan, 1997). The combination has resulted in a definite change in how we conduct our evaluations and what we are looking for. Research using effective school data pointed us towards a better internal understanding of how schools work, how to structure school accountability, and what to look at when evaluating a school. Research on effective teachers has led to changes in how we view classrooms and programs and the types of changes needed to improve schools. The resulting approach we now use and why we use it is described in this paper.

Approach to School Evaluation

We have adopted an approach to school evaluation with three components:

- School and Teacher Accountability
- Compliance Assessment and Evaluation.
- Training and Service Activities

The change in perspective hinges around school and teacher accountability. It is accountability measured in two ways with value-added and with unadjusted measures. The actions of effective schools and teachers comprise the focus of the evaluation efforts.

The identification of effective schools and teachers drives the evaluation activities chosen. Compliance activities whenever possible are used to provide additional information about school and teacher accountability and the variables underlying them. Similarly, wherever possible, training and service activities are focused on presenting two types of information to policy makers, administrators, campus leaders and teachers and helping them understand it. The first is how schools and classes are performing on accountability measures and the second is what variables are related to effective and ineffective classes and schools.

Without reading further, one may well ask at this point whether we intend to pose new models of school evaluation. The answer is obviously no. We still use the appropriate parts of the classic models proposed by Stufflebeam, Scriven, Stake, and other founding fathers (See Madaus, Scriven, and Stufflebeam, 1983 for an excellent exposition of the major models). When one must evaluate the implementation of reading programs as part of Title 1 compliance for example, the existing models still apply and we apply them very well (most of the time). Further, we are somewhat open to expanding our horizons and considering different approaches and models and perturbations of the existing models. Our point, though, is that the latest research on school and teacher effects is requiring us to take a different perspective on what we are looking for and how it relates to the ultimate goal of improving schools. Our conceptualization of components of school evaluation is based on our need to assure a focus with the highest possible payoff in terms of improving our schools. To the extent we conduct any activity, our first concern is whether the activity can be used to make schools more effective. With this in mind, let us consider each of these three components briefly.

Compliance. Compliance activities are planned to meet the information requirements of governmental agencies, courts, the Board of Education, grantors, and others. These activities fall under the heading of information we are legally (and sometimes morally) required to provide as a matter of a federal law, a state law, a grant requirement, a Board policy, or other similar obligation. These include a wide range of requirements. They range from providing counts and numbers to providing evaluations of program effectiveness. Sometimes these requirements are the equivalent of evaluation "busy work", have no direct relationship to school improvement, and are simply provided because someone or some organization thought that it was important to have this information. Sometimes the information is critical to both the district's students and the agencies requesting the information, but does not bear on increasing effectiveness. A good example of this would be steps taken to verify that teachers have provided an individual instructional improvement plan for each special education student. It is important information for all concerned but is not directly related to school effectiveness.

However, most compliance information fits directly or indirectly into our focus on effective schools. No doubt this is directly related to the fact that many of the regulations by agencies, etc. deal with school effectiveness. For example, the evaluation requirement for Title 1 funds asks for information about program effectiveness. Some is demanded as required information and some allows us the latitude to examine program effects in light of our own needs.

Examples of current compliance activities and requests include (but are not limited to):

- Counts of Title 1 students served and compliance with Title 1 guidelines
- Evaluation of the effectiveness of Title 1 schools and programs

- Counts of students and personnel meeting requirements set by the Federal Court in desegregation rulings
- Evaluation of the effectiveness of Title 1, reading improvement, longitudinal achievement, learning centers, magnet schools, and multilingual instructional programs for the Federal Court as a part of desegregation rulings
- Delineation of the effects of boundary changes on school ethnic distributions as input to Federal Court approval of such changes
- Compliance of the District to Federal regulations relative to providing services to special education students
- Information for Office of Civil Rights complaint investigations
- Evaluation of effectiveness of numerous programs funded by granting agencies including the State Education Agency, NSF, and private foundations
- Compliance with Board policies by the District and schools in the areas of testing students, hiring personnel, placement of personnel, placement of students in programs, appraisal of personnel, assessment of school effectiveness, evaluation of Board Approved projects, evaluation of internal charter schools, etc.
- Compliance of schools with Federal and State program guidelines, particularly in the area of appropriate documentation, expenditures, and student services provided
- Development and implementation of school and district improvement planning systems focused on school and district effectiveness

As is readily apparent, compliance activities cross a broad range of tasks and responsibilities.

Accountability - Value-Added School and Teacher Effectiveness. Our epiphany came with our venture into measuring school and teacher effectiveness. In 1991 we were ordered by the Board to measure the effectiveness of schools using a value-added model that controlled for major variables outside the control of a school with the first indices to be released in 1992. (Webster, Mendro, Orsak, and Weerasinghe, 1997) The effectiveness measures were to be the base of a school awards program and were also to be used to identify ineffective schools with the intent of helping school leadership and turning around the progress of the school.

From the beginning, we envisioned being able to use the indices for more than the identification of effective schools. We had in mind doing observations in effective schools to identify elements of climate and organizational structure that could then be transferred or disseminated to other schools. Also, we wished to look at classrooms in the effective and ineffective schools. Our first forays into this began in 1994 after we had identified a number of elementary schools that were relatively effective or ineffective and had retained the same principal during that time frame. (Webster, Mendro, Bearden, Bembry, and Jordan, 1997).

Immediately our preconceptions were whisked away. The schools, effective or not, had different climates, different management styles on the part of the principals, and different approaches to their students. However, as a group, the effective schools shared some traits that were absent or less pronounced in the ineffective schools. In particular, four things were apparent:

- *Learning-centered focus.* Each of the effective schools made it clear to students and staff that the focus of the school was student learning and that all other school elements were secondary to this purpose. This focus was not clear or not present in the ineffective schools.
- *Student expectations.* There were both effective and ineffective schools in poor neighborhoods. Both types were aware of the challenges and problems their students faced. Both types stated that they believed all students could learn. However, ineffective schools noted that they could not expect their students to learn under these circumstances while effective schools made it clear that students were expected to learn regardless of their circumstances.
- *Quality of teaching.* Our observers saw some excellent teachers in very ineffective schools. Conversely, they saw very few poor teachers in effective schools.
- *Demanding effective teaching.* Principals of effective schools expressed much more willingness to confront and either change the behavior of or force out ineffective teachers.

The surprising elements were that outside of these four items, none of the other traditional elements associated with effective schools seemed to be necessary. As noted, principal's management style and general school climate differed greatly from school to school. Also, the idea of achieving effectiveness by forcing out weak teachers was new to our preconceptions at that time.

At about this point the Board directed us to become involved in measuring the effectiveness of classrooms with the intent of reforming the teacher appraisal process to include effectiveness information. Our first full set of indices at grades 1-8 was prepared in 1994-95. Upon report of this, the Board created a teacher appraisal task force and demanded a revised appraisal system. After a trial in 1995-96, the Board approved a teacher appraisal system in June 1996 incorporating the indices as input to the process (Bembry, 1996; Bembry, Bearden, and Mendro, 1997).

As obvious as it sounds, we were intrigued by the idea that the difference in schools may be more related to the effectiveness of the collection of teachers assembled rather than to other factors. About this time, we ran across the seminal work of Sanders and Rivers (1996) looking at longitudinal effects of teachers on students. We immediately set about replicating it with our effectiveness data. The result was two studies that confirmed Sanders and Rivers and extended their results across tests and grades. (Jordan, Mendro, and Weerasinghe, 1997; Bembry, et al, 1998).

The import of Sanders and Rivers study and the two in which we replicated and extended Sanders and Rivers work is clear with a simple recounting of major results. Teachers in all three studies were divided into 5 levels of effectiveness for each of 3 years. Then

students were followed based on the sequence of teachers by level of effectiveness with 125 combinations being possible. In our second study (Bembry, et al., 1998) teachers were also divided into three groups for each of 4 years. The data in all three studies were analyzed in two ways: using the student' standardized test score for the year prior to each sequence as a covariate and using either analysis of covariance or HLM to analyze effects for each level and student group or selecting groups with approximately equal pretest scores and different levels of teacher effectiveness across three years and examining raw scores across groups. Major results were:

- Teacher main effects were exceedingly large with virtually no interaction effects.
- The effects of a poor teacher could be detected clearly 2 and 3 years after students were in that teacher's class. When students with one poor teacher and 2 or 3 excellent teachers were compared to students with all excellent teachers, effects of the poor teacher were identifiable.
- When we divided teachers into five groups the difference between the scores of approximately equal pretest groups after 1 year of a poor teacher *vs.* 1 year of an excellent teacher ranged from 25 to 35 percentile points. The difference between 3 years of excellent or poor teaching resulted in approximately 40 to 50 points on the percentile scale.
- When examining effect sizes, excellent teachers almost invariably could not undo the effects of a poor teacher in one year.
- Results at comparable grades were almost identical in the Sanders and Rivers study and the Dallas studies. These were obtained despite different ways of computing classroom effectiveness, different outcome measures, different student populations and characteristics (Memphis and Nashville *vs.* Dallas), and different methods of analyzing the data.
- Additionally, the Dallas studies found a systematic bias in student assignment across years. As years progressed, students with lower achievement began to be systematically assigned to teachers with prior lower effectiveness scores and vice-versa. (Sanders and Rivers did not address bias.)

In essence, teachers have massive differences in effects on student achievement, contrary to myth the effects of poor teachers can rarely be substantially modified in a single year, and, in the Dallas data, the poorer the achievement levels of a student, the less likely the student would be to have the opportunity of being paired with a more effective teacher.

As noted, we have found few effects with programs implemented on any type of large scale in the past. A review of the most common reading interventions done by the representatives of the Texas Instruments Foundation working with the Dallas Public Schools showed that the expectation of academic growth from an intervention is at best less than half of that demonstrated for a top effectiveness level teacher (Dosher and Fischer, 1998). Further, the research literature does not address the issue of how well intervention programs would work when less effective teachers volunteered or were instructed to implement them.

One of the first thoughts that struck us when we had completed our first round of research into teacher effects (Jordan, Mendro, and Weerasinghe, 1997) was the implication these results had in the realm of policy. This prompted the second round cited above (Bembry, et. al., 1998). As noted in the latter paper, knowledge of teacher effectiveness has implications in terms of:

- *Student Equity*. Students assigned to ineffective teachers are denied the opportunity to learn at maximum potential. Where a bias exists in assignment of students, lower achieving students are systematically denied the benefits of effective teachers.
- *Campus Organization and Teacher Assignment*. Closely aligned to student equity are the assignment of students to teachers and the subsequent effects of assignment on organizing the instructional program on a campus and devising teacher assignments. Consider two examples. First, while our research shows that more effective teachers cannot fully remedy the effects of less effective teachers in one year, they have a far better effect on student achievement than assigning such a student to an ineffective teacher for a second year. Thus, a school should consider the overall student assignment pattern in light of a longitudinal pattern of student/teacher effectiveness assignments. The second example addresses current grouping patterns. Should the most effective teachers be offered larger classes (with commensurately increased salary) with less effective teachers acting as assistants? Such a pattern may well offer the best alternative to student assignment by effectiveness.
- *Teacher Training and Retention*. The research results suggest policy changes in how staff development is offered and its content structured. The content

should be based on research into what effective teachers do and how they prepare to do it. Staff development needs to be tailored to the needs of the teacher with classroom effectiveness as an organizing variable.

More to the point of our prior discussion of the focus of school evaluation, classroom effectiveness has implications on the conduct of project and school evaluation. As noted, we cannot yet address the issue of the effectiveness of interventions with less effective teachers. However, research, which we completed in 1998, does address the issue of the size of annual gains for teachers at various levels of effectiveness. (Mendro, et. al., 1998a). In this study using the same initial data from Bembry et. al. (1998), we examined teachers divided into 5 groups and 3 groups depending on the number of years in the longitudinal study (three or four). We then looked at mean NCE data for reading and mathematics within student pretest quartile for teachers in each of the 3 or 5 groups. Results indicated that for the 3-group division of teachers, differences within student quartile groups from teachers in the lowest group to teachers in the highest group ranged from 7.5 to 11.7 mean NCEs. (This represents an approximate difference of 10 to 15 percentile points at the NCE means depending on where in the distribution it is calculated.) For the five-group division of teachers, because of the finer groupings, differences range from 9.3 to 13.8 mean NCEs. (Representing an approximate difference of 13 to 18 percentile points at the NCE means.)

These results have clear implications for conducting the evaluation of schools and interventions implemented in schools. Certainly, the effects of any intervention must be interpreted in light of the prior effectiveness of the teacher. It is easy, given the differences in effectiveness noted above, to perceive how an effective treatment might be thought useless with the wrong teachers selected in implementation and comparison

groups. Further, ineffective treatments can be deemed effective when in the hands of effective teachers. At a minimum, effectiveness must be known beforehand and examined to assure that results are not distorted.

For us one of the disturbing implications of the size and extent of the differences is that we calculated them by the simple expedient of dividing our population into three or five equal parts. Approximately 1,500 teachers were involved in our two research studies. This implies that, across reading and mathematics for the 3 or 4 grades we examined, a minimum of 300 of the 1,500 teachers were performing at a very poor level.

One further implication of the results of our longitudinal effectiveness research is the implied change in conducting school evaluation. We now focus much of evaluation on determining characteristics of effective and ineffective teachers. Just one example of the payoff of this approach comes from our studies of our mathematics program. We studied the elementary mathematics program in 1997 which has provided insights into general qualities of effective math teachers (Bearden, 1997); followed this by an evaluation of the first year algebra program in 1998 which gave us much useful information about effectiveness and teacher assignment at the secondary level (Bearden, 1998); examined middle school mathematics as part of our court-ordered evaluation of learning centers which gave us a beginning look at the role of level of content in mathematics effectiveness (Weir, 1999); used that study to currently conduct a general analysis of the middle school mathematics program which is shedding light on the effects of the same teachers across different levels of course material (Weir, 2000; Weerasinghe, 2000); and, currently, have completed process evaluation of the elementary mathematics program following up the first evaluation in considerably more detail and providing us with a

richer look at effective elementary mathematics teachers (Heiry, 2000). The evaluation of the algebra program, however, contains a caution about the use of value added indices. One of the results of the evaluation was the determination that algebra courses generally are taught by less effective teachers across the board. As a result, our relative indices identified the most and least effective teachers of a generally lower-level group. This provides a lead-in to the discussion of the second necessity of a school accountability system: the need for unadjusted measures of achievement.

Accountability - Unadjusted Measures of District, School, and Classroom Effectiveness. Effectiveness indices are still relative measures. In terms of measuring a population of students or teachers, they are exceptionally good measures for determining relative effectiveness, i.e., they identify the best, worst, and middle of a group. If the entire group makes progress or if the entire group fails to move, the indices will still separate the best and the worst. Thus, there is a need for unadjusted measures that answer questions such as what was the progress of all students without adjustment to the data? However, the down side of unadjusted measures is that, as unadjusted measures, they are biased with strong correlations favoring higher income students, language proficient students, and female students at a minimum. (Webster, Mendro, Bembry, and Orsak, 1995).

We report the following unadjusted types of data with each major test:

- *Cross-sectional analyses.* These report results for each student who participates in any given testing regardless of the length of time they were enrolled or when they moved into a given school. These can be reported for any number of years in succession. They answer questions such as how did this year's third grade students compare overall to last year's third grade students.
- *Cohort analyses.* These report results for each student who has participated in each of a given number of testings and has complete data available. For example, we routinely report two-year cohorts with the results of a prior year

and a current year and multi-year cohorts that track students who enter our system and have test results available from first grade to their current grade. These answer questions such as how have we done with students where we know their entry level of performance and where they have been with us for a given number of years?

- *Quasi-cohort analyses*. These report results from the cross-sectional analyses where groups of students are followed from grade to grade, but there is no restriction on complete test data. For example, a three-year quasi-cohort for grade three includes the results of all students tested at grade 1 two years ago, all students tested at grade two last year and all students tested at grade three this year.

As noted these results all have considerable degrees of bias and should be used with great caution.

Achievement Goals. One constant demand of the accountability system is absolute achievement goals that are also fair. A hybrid comprised of information from indices and unadjusted measures provides a ready answer to this demand. We are in the midst of developing a system that sets absolute goals based on the performance of more effective teachers with similar students. The system is based on the premise that we define five levels of effectiveness and develop goals based on the performance of the three highest groups of teachers. Goals are based on the performance within categories of students using indices data to identify effective teachers. These goals then are both attainable (more than half of current teachers already meet them) and fair (bias has been controlled in the indices). Single goals can be set or levels of goals can be set. The system is fair to the greatest extent allowable and it is flexible. If this year's set of goals is exceeded by most teachers, a new set of goals can be developed based on the performance of the current top three groups. The system avoids sliding since, once established, a set of goals does not have to be adjusted downward because we have evidence that a significant portion of teachers already met them.

BEST COPY AVAILABLE

Goal development based on indices is definitely preferred over arbitrary absolute goals such as those used by the State of Texas. The State sets a passing rate on a skills test and then requires a set percentage of students to pass the test for a school to be given an accountability rating of acceptable. At the current time the passing rate is low and definitely every school should meet the acceptable criterion. (In fact, the low passing rate is causing more problems by encouraging low-level teaching than it is solving.) However, as with any unadjusted measure of achievement, the results are highly correlated with socioeconomic status and language proficiency.

Training and Service Activities. The last major component of the system of school evaluation is providing training and service activities to the schools and to the administration based on the components of school effectiveness. The accountability system and its information must be thoroughly explained to all constituents, used in identifying the parameters of school-level research or school evaluation activities, and influence staff development and other functions of a school. In addition, schools must be supported toward success within the accountability system through specific training in data analysis, planning, and in the development of easily read student-level data reports for teachers to use in adjusting instruction. Finally, the administration must be aware of the import and limitations of data provided and helped in using it correctly and responsibly.

It is imperative that administrators, teachers, central office staff and community members understand the accountability system. Developing a fair system of assessment for an extremely diverse student population requires a level of statistical sophistication that is often not part of prior training for teachers and administrators. In addition, schools that

have won awards in other areas of school life (i.e. UIL) often have community members and parents vitally interested in good reports from all assessments. A lack of understanding leads to the condemnation of the system or at least a perception of unfairness. Lack of awareness of the limitations of the system on the part of central office administrators can lead to unconcern for the plight of schools or teachers. Training, then, is critical in establishing not only an understanding of but also a commitment to any accountability system. This training should always include the measures within the system, their types (norm-referenced, unadjusted measure, adjusted growth measure, non-test variable, etc), how each function in the system and how the results will be used. Training should also include those who had and will have input into the system for obvious reasons.

Training on simply understanding the system, however, is insufficient. By both personnel and program evaluation standards, support for success within an accountability system is also required. Both teachers and administrators need training and additional on-campus assistance in data interpretation and identifying steps for improving instruction. This adds a "data interpretation and planning" component that was not previously present in our school evaluation system. Most school evaluation stops at reporting status, either of the school or of the teacher. In order for a research and evaluation division to be most effective, there needs to be a shift from simply reporting a school's status to supporting that school by helping teachers and administrators make the connections among planning, instruction and assessment.

The Dallas Research and Evaluation Department issues a planning packet to each campus in July that includes grade level by objective reports for both the norm-referenced and the

State skills test. Campus level summary reports are also included, as is a contextual data report. A school's contextual data include information for students considered at-risk, student and teacher attendance, and a school's pregnancy, truancy, and suicide rates. It also lists the number and type of discipline offenses for the school.

Teachers and administrators use this information in developing both the Campus Improvement Plan and individual teachers' Instructional Improvement Plan. Research and Evaluation staff members conduct training for teachers and administrators on the data packet during the summer. Staff members are also available to meet with individual campus staff members.

A second report of student-level test information is sent to teachers when they return each year. Class rosters of a teacher's current students' test information are sent to the campuses at the beginning of each school year. These rosters allow teachers to identify possible low-performing students and plan for intervention.

The third report of student-level achievement information is included in the Classroom Effectiveness Indices that are sent to the schools by mid-September. This report includes a growth chart that indicates the level of growth for each student included in the index. Teachers whose students do not consistently show growth are required to include the CEI information on their individual Instructional Improvement Plan. Again, training is conducted for both teachers and administrators on the Classroom Effectiveness Indices each year.

Information from the accountability system used in evaluating individual schools and classrooms also needs to be used in school wide evaluation activities and/or in program

evaluation. As discussed above, program evaluation is a misnomer unless we begin with the knowledge of and the evaluation of the effectiveness of personnel in the program. Teachers' effectiveness, once established, will need to be taken into account as teachers are identified for any classroom observations. As noted, Dallas is currently investigating the differing characteristics of classrooms of high CEI teachers and low CEI teachers in math, reading, and science. These results of program evaluations will be included in staff development for both teachers and administrators. Grouping teachers observed by their CEIs also allows assistance in the analysis of results.

Classroom Effectiveness Indices have also been consulted to identify teachers for summer school, to identify teachers to participate in textbook selection, and to select trainers for staff development. These central administrative functions are a necessary result of the effectiveness studies.

Summary

This paper has shown the shift in emphasis in school evaluation in the Dallas Public Schools. The primary emphasis is now on accountability in measuring, using, and learning from teacher and school effectiveness. Both value-added and unadjusted measures are needed. Traditional program evaluation has been repositioned to provide light on teacher and school effectiveness, wherever possible. While still responsible for many compliance activities, these activities have also been centered on effectiveness information. As a concomitant of the effectiveness studies, training and service activities have taken a major role in the Department of Research and Evaluation. Where they were

primarily perfunctory before, they are used to expand knowledge of the accountability system, knowledge of effective practices, and help administrators make better decisions.

References

Bearden, D. K. (1997). *An Overview of the Elementary Mathematics Program.* Evaluation Report REIS97-116-2. Dallas: Dallas Public Schools.

Bearden, D. K. (1998). *An Evaluation of the Secondary Mathematics Program.* Evaluation Report REIS98-116-2. Dallas: Dallas Public Schools.

Bembry, K. L. (1996). *Teacher Evaluation System Development and Field Test.* Evaluation Report REIS96-120-2. Dallas TX, Dallas Public Schools.

Bembry, K. L., Bearden, D. K., and Mendro, R. L. (1997). *Using Student Achievement in Teacher Appraisal.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, and March 24-28.

Bembry, K. L., Jordan, H., Gomez, E., Anderson, M., and Mendro, R. L. (1998). *Policy Implications of Long-Term Teacher Effects on Student Achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, April 1998.

Commission for Educational Excellence. (1991). *Final Report.* Dallas TX, Dallas Public Schools.

Dallas Public Schools, (1982) *Research and Evaluation Final Reports, 1980-81.* Dallas TX, Dallas Public Schools.

Dallas Public Schools, (1987) *Executive Summaries of Research and Evaluation Reports, 1986-87.* Dallas TX, Dallas Public Schools.

Dosher, R. and Fischer, J. (1998). *Academic Improvement Management.* Dallas, Texas, Texas Instruments Foundation.

Heiry, T. (2000). *Preliminary Process Findings for Elementary (Grade 3) Mathematics.* Interim Process Evaluation Report. Dallas: Dallas Public Schools.

Jordan, H., Mendro, R., and Weerasinghe, D. (1997). *Teacher Effects on Longitudinal Student Achievement.* A paper presented at the CREATE annual meeting, Indianapolis, IN, July 1997.

Madaus, G., Scriven, M, and Stufflebeam, D. (1983). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation.* Hingham, MA: Kluwer Academic Publishers.

Mendro, R. L., Webster, W. J., Bembry, K. L., and Orsak, T. H. (1995). *An Application of Hierarchical Linear Modeling in Determining School Effectiveness.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April 17-21.

Mendro, R. L., Jordan, H., Gomez, E., Anderson, M., and Bembry, K. (1998a). *An Application of Multiple Linear Regression in Determining Longitudinal Teacher Effectiveness.* Paper presented at the 1998 Annual Meeting of the American Educational Research Association, San Diego, CA, April 1998.

Mendro, R. L., Jordan, H., Gomez, E., Anderson, M., and Bembry, K. (1998b). *Longitudinal Teacher Effects on Student Achievement and Their Relation to School and Project Evaluation.* Paper presented at the 1998 Annual Meeting of the American Educational Research Association, San Diego, CA, April 1998.

Sanders, W. L. and Rivers, J. C. (1996). *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement.* Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Webster, W. J. (1995). The Connection Between Personnel Evaluation and School Evaluation. *Studies in Educational Evaluation*, 21, 227-254.

Webster, W. J. and Mendro, R. L. (1997). The Dallas Value-Added Accountability System in *Grading Teachers, Grading Schools*, (Ed. by J. Millman) Newbury Park, CA, Sage Publications.

Webster, W. J., Mendro, R. L., and Almaguer, T. O. (1993). *Effectiveness Indices: The Major Component of an Equitable Accountability System*, ERIC TM 019 913.

Webster, W. J., Mendro, R. L., Bearden, D. K., Bembry, K. L., and Jordan, H. R. (1997). *Rewarding Effective Schools - Theory and Practice in an Outstanding Schools Awards Program.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, March 24-28.

Webster, W. J., Mendro, R. L., Bembry, K. L., and Orsak, T. H. (1995). *Alternative Methodologies for Identifying Effective Schools.* Paper presented in a Distinguished Paper Session at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April 17-21. ERIC EA 027189.

Webster, W. J., Mendro, R. L, Orsak, T. H., and Weerasinghe, D. (1997). *A Comparison of the Results Produced by Selected Regression and Hierarchical Linear Models in the Estimation of School and Teacher Effects.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, March 24-28.

Webster, W. J. and Olson, G. H. (1988). A Quantitative Procedure for the Identification of Effective Schools. *Journal of Experimental Education*, 56, 213-219.

Weerasinghe, D. (2000). *CEI Validation Studies.* Interim Process Evaluation Report. Dallas: Dallas Public Schools..

Weir, D. (1999). *Evaluation of the Learning Center Program, 1998-99.* Evaluation Report REIS99-208-2. Dallas: Dallas Public Schools.

Weir, D. (2000). *Preliminary Process Findings for Secondary Mathematics.* Interim Process Evaluation Report. Dallas: Dallas Public Schools.